HadoopDevelopmentToolsProposal

HDT (Hadoop Development Tools)

Abstract

Tools to support developing applications that use Apache Hadoop from within Eclipse.

Proposal

Hadoop Development Tools are a set of extensions to Eclipse providing support for creating, launching and debugging distributed applications, as well as interacting with HDFS filesystems. This work will build on the existing Map Reduce Tools present in the Apache Hadoop project.

Background

Map Reduce Tools have existed as part of contrib for Apache Hadoop. Unfortunately they are source tied to a single version of Hadoop, and development has stalled, with little movement past the Hadoop 0.20 line.

Rationale

Support for newer versions of Hadoop from within Eclipse is regularly raised on the Hadoop mailing lists, so there is a clear need to drive these tools forward. Development tools generally are worked on separate from the target tools/platform, separating the tools out will allow for supporting multiple versions, so a developer could work with a heterogeneous environment.

Initial Goals

- Give the tools project a home of its own.
- Port current MapReduce tools feature set to all current release lines of Hadoop in a single Eclipse install.
- Documentation and tutorials for all features.
- Publish Eclipse update site, and join Eclipse marketplace listing.
- Establish release cycle that combines support for Hadoop and Eclipse release cycles.
- Look to build support for YARN, MRUnit and possibly other Hadoop-related projects.

Current Status

The source for the current MapReduceTools lives in the contrib section of the Hadoop source. In its current implementation it is tied to the version of Hadoop against which it is compiled. The layout and API that it was developed with means that it can only be used with the 0.20 or 1.0 Hadoop releases, the new layout and YARN api introduced with the 0.23 and 2.0 lines are not supported.

Meritocracy

Several people and companies have already expressed an interest in contributing to this project, and we hope to attract additional interest during the proposal discussion. We plan to invest and support a meritocracy that attracts, invites, and supports newcomers to build a vibrant and diverse community.

Community

The target community is developers who are working developing Map/Reduce applications against Hadoop. Given the success of Hadoop the target group is likely to be quite large. Separation from the Hadoop community would make it easier to support multiple versions of hadoop, as well as merging the release cycles of Hadoop and Eclipse to provide predictable iteration and improvement in the toolset.

Core Developers

The initial list of developers includes people experienced with Hadoop and developing against the Eclipse platform.

- Adam Berry (amberry at yahoo-inc dot com)
- Jeffrey Zemerick (jeffrrey at mtnfog dot com)
- Evert Lammerts (Evert dot Lammerts at sara dot nl)
- Simone Gianni (simoneg at apache dot org)

Alignment

Hadoop Development Tools aligns with both Hadoop and Eclipse. Hadoop as the platform for the development target, and Eclipse as the IDE platform used as the base for the tools.

Known Risks

Orphaned Products

Inexperience with Open Source

The committers have experience with Apache and Eclipse open source development.

Reliance on Salaried Developers

Hadoop Development Tools will be developed with a mix of salaried and volunteer time.

Relationships with Other Apache Projects

Hadoop Development Tools is closely related to Apache Hadoop.

An Excessive Fascination with the Apache Brand

Given the success of Hadoop and associated projects, Apache is the natural place for the Hadoop Development Tools. Chris Mattman suggested the Apache Incubator as appropriate on the Hadoop general mailing list following the success that MRUnit had taking the path from Hadoop contrib to an Apache top level project.

Documentation

Documentation for the current tools can be found at http://wiki.apache.org/hadoop/EclipsePlugIn

Initial Source

http://svn.apache.org/repos/asf/hadoop/common/trunk/hadoop-mapreduce-project/src/contrib/eclipse-plugin/

Source and Intellectual Property Submission Plan

The source, and any suggested initial patches, are already hosted either in Apache's Subversion or JIRA.

External Dependencies

Eclipse Platform Eclipse Java Development Tools

Cryptography

Hadoop Development Tools likely does not fall into this area.

Required Resources

Mailing lists

- hdt-dev
- hdt-commits
- hdt-user

Git Repository

- git://git.apache.org/incubator-hdt.git

Issue Tracking

- JIRA Hadoop Development Tools (HDT)

Other Resources

- Jenkins/Hudson for builds and test running.

Initial Committers

- Adam Berry (amberry at yahoo-inc dot com)

- Jeffrey Zemerick (jeffrrey at mtnfog dot com)
- Evert Lammerts (Evert dot Lammerts at sara dot nl)
- Simone Gianni (simoneg at apache dot org)

Affiliations

- Adam Berry - Yahoo!
- Jeffrey Zemerick - Mountain Fog
- Evert Lammerts - SARA
- Simone Gianni - n/a

Sponsors

Champion

- Roman Shaposhnik

Nominated Mentors

- Chris Douglas
- Chris Mattman
- Roman Shaposhnik
- Suresh Marru

Sponsoring Entity

Incubator PMC